neural network
= nonlinear function approximator

Once we fix the hyperparameters the problem depends only on the parameters (weights)

How to find weights?

- Regression:

$$w = \arg\min_w E(w)$$

$$\downarrow \arg\min_w \sum_{n=1}^{N} (t_n - g(x_n, w))^2$$

Through MLE we design error functions.
Through gradient descent we minimize error functions.

Remember? multi-layer classification:
it can have inside the ReLU, the last is SOFTMAX